AN ENHANCED CONCEPT BASED APPROACH FOR MEDICAL INFORMATION RETRIEVAL TO ADDRESS READABILITY, VOCABULARY AND PRESENTATION ISSUES

IBRAHIM UMAR KONTAGORA

A thesis submitted in
fulfillment of the requirement for the award of the
Doctor of Philosophy

Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia

OCTOBER, 2019

DEDICATION

This thesis is dedicated to almighty Allah, my parents and the entire UMAR MD KONTAGORA Family.

ACKNOWLEDGEMENT

In the Name of Allah Most Merciful, Most Beneficent in Mercy, all praise is due to Allah, peace and blessings of Allah be upon our leader, the messenger of Allah, Prophet Muhammad (S.A.W) and upon all his Family and his companions**.**

I express my profound gratitude and indebtedness to my supervisor Ts Dr. Isredza Rahmi Binti A. Hamid, for her relentless, intellectual, constructive and valuable suggestions and advices throughout the duration of this research work. I also want to thank her for her motivation and encouragement towards achieving the set objectives of the research title topic. Ma, am indeed very grateful, may almighty Allah reward you abundantly, amin.

My appreciations also goes to my co – supervisor Dr. Nurul Aswa Omar for her encouragement, valuable suggestions, motivation and advices towards achieving the set objectives of the research study. I am indeed very graceful for all your patience and inputs towards the actualization of this research work. My appreciations also goes to my examiners, commentators, FSKTM academic and non-academic staffs, who have contributed in one way or the other towards the successful completion of this research work.

My special gratitude goes to my immediate family Saudatu Buhari (Wife), Bilkisu Ibrahim Kontagora (Daughter) and Ahmed Muhammad Ajiya (Son), parents, brothers and sisters for all their patience and prayers throughout my academic pursuits. Also worthy of indebtedness are my employer Niger State Polytechnic, Zungeru and the national funding body TETFund for financing and granting me the opportunity to embark on the study over here in Malaysia.

I am also thankful to CGS Biasiswazah for providing me with scholarships at two separate semesters, all my friends and faculty members FSKTM post graduates society, who have provided assistance at various occasions. I am indeed grateful. May almighty Allah reward you all abundantly, amin.

ABSTRACT

Querying of health information retrieval for health advice has now become a general and notable task performed by individuals on the Internet. However, the failure of the existing approaches to integrate program modules that would address the information needs of all categories of end-users remains. This study focused on proposing an improved framework and designing an enhanced concept based approach (ECBA) for medical information retrieval that would better address readability, vocabulary mismatched and presentation issues by generating medical discharge documents and medical search queries results in both medical expert and layman's forms. Three special program modules were designed and integrated in the enhanced concept based approach namely: medical terms control module, vocabulary controlled module and readability module to specifically address the information needs of both medical experts and laymen end-users. Eight benched marked datasets namely: Medline, UMLS, MeSH, Metamap, Metathesaurus, Diagnosia 7, Khresmoi Project 6 and Genetic Home Reference were used in validating the systems performance. Additionally, the ECBA was compared using three existing approaches such as concept based approach (CBA), query likelihood model (QLM) and latent semantic indexing (LSI). The evaluation was conducted using the performance and statistical metrics: P@40, NDCG@40, MAP, Analysis of Variance (ANOVA) and Turkey HSD Tests. The outcome of the final experimental results obtained shows that, the ECBA consistently obtained above 93% accuracy rate results on Medline, UMLS and MeSH Datasets, 92% on Metamap, Metathesaurus and Diagnosia 7 datasets and 91% on Khresmoi Project 6 and Genetic Home Reference datasets. Also, the statistical analysis performance results obtained by each of the four approaches: ECBA, CBA, QLM and LSI shows that, there is a significant difference among their Mean Scores, hence, the null hypothesis of no significant difference was rejected.

ABSTRAK

Permohonan bagi maklumat berkenaan kesihatan bagi tujuan nasihat kesihatan kini telah menjadi tugas umum dan penting yang dilakukan oleh individu di Internet. Pun begitu, sistem sedia ada masih tidak berupaya untuk hadir dengan modul program yang diperlukan bagi menangani masalah keperluan maklumat bagi semua tahap pengguna akhir. Kajian ini memberi tumpuan terhadap sebuah rangka kerja yang ditambah baik dan mereka bentuk sebuah pendekatan berasaskan konsep yang dipertingkatkan (ECBA) bagi membaca semula maklumat perubatan yang akan menyelesaikan masalah kebolehbacaan, perbendaharaan kata yang tidak sepadan dan isu persembahan dengan cara menjana dukumen pelepasan perubatan dan hasil pertanyaan carian bagi kedua-dua pihak– pakar perubatan dan orang awam. Tiga modul program khas direka dan disepadukan ke dalam pendekatan berasaskan konsep yang dipertingkatkan yang dikenali sebagai: modul kawalan terminologi medik, modul kawalan perbendaharaan kata dan modul kebolehbacaan; khusus bagi menangani keperluan maklumat kedua-dua pihak yakni pakar perubatan dan pengguna awam. Lapan dataset penanda aras iaitu: Medline, UMLS, MeSH, Metamap, Metathesaurus, Diagnosia 7, Khresmoi Project 6 dan Genetic Home Reference telah digunakan bagi mengesahkan prestasi sistem. Di samping itu, ECBA telah dibandingkan dengan tiga pendekatan yang sedia ada seperti pendekatan berasaskan konsep (CBA), model kebarangkalian pertanyaan (QLM) dan pengindeksan semantik laten (LSI). Penilaian telah dilaksanakan menggunakan matrik prestasi dan statistik: P@40, NDCG@40, MAP, Analisis Varians (ANOVA) dan Pengujian Turkey HSD. Keputusan akhir eksperimen yang diperoleh menunjukkan bahawa ECBA telah memperoleh kadar ketepatan melebihi 93% secara konsisten menggunakan dataset-dataset Medline, UMLS dan MeSH; kadar ketepatan sebanyak 92% pada dataset-dataset Metamap, Metathesaurus dan Diagnosia 7; dan kadar ketepatan sebanyak 91% pada set data Khresmoi Project 6 dan Genetic Home Reference. Selain itu, hasil prestasi analisis statistik yang diperoleh bagi setiap satu daripada empat pendekatan tersebut: ECBA, CBA, QLM dan LSI menunjukkan bahawa terdapat perbezaan yang signifikan pada skor purata, justeru, hipotesis tidak sah yang menyatakan bahawa tiada perbezaan ketara telah ditolak.

TABLE OF CONTENTS

TITLE		**i**
DECLARATION		**ii**
DEDICATION		**iii**
ACKNOWLEDGEMENT		**iv**
ABSTRACT		**v**
ABSTRAK		**vi**
TABLE OF CONTENTS		**vii**
LIST OF TABLES		**xii**
LIST OF FIGURES		**xvii**
LIST OF ALGORITHMS		**xix**
LIST OF SYMBOLS AND ABBREVIATIONS		**xx**
LIST OF APPENDICES		**xxiii**
LIST OF PUBLICATIONS		**xxiv**
CHAPTER 1 INTRODUCTION		**1**
1.1	Research Background of the Study	1
1.2	Problem Statement of the Study	4
1.3	Objectives of the Study	6
1.4	Scope of the Study	7
1.5	Significance of the Study	8

1.6 Thesis Organization 9
1.6.1 Chapter 1- Introduction 9
1.6.2 Chapter 2- Literature Review 9
1.6.3 Chapter 3- Methodology 9
1.6.4 Proposed Enhanced Concept Based Approach 10
1.6.5 Chapter 5- Simulation Results and Discussions 10
1.6.6 Chapter 6- Conclusion and Future Work 10

CHAPTER 2 LITERATURE REVIEW 12
2.1 Introduction 12
2.2 An Overview of the Three Existing Approaches Used in Evaluating The Enhanced Concept Based Approach (ECBA) 13
2.2.1 The Concept Based Approach (CBA) 13
2.2.2 The Query Likelihood Model (QLM) 15
2.2.3 The Latent Semantic Indexing (LSI) Model 18
2.3 Discharge Summaries Reports 19
2.4 Clinical Query Reports 20
2.5 Reviewed Related Works 20
2.5.1 Information Retrieval Algorithms 21
2.5.2 Features Used 24
2.5.3 Information Retrieval Models 28
2.6 Some Summarized Related Works 33
2.7 Chapter Summary 40

CHAPTER 3 METHODOLOGY 41
3.1 Introduction 41
3.2 Research Process (Flowchart) 42
3.2.1 Phase-I: Initial and related works 44
3.2.2 Phase-2: Data preparation and cleaning 44
3.2.3 Phase-3: Design of the Improved Framework 49
3.2.4 Phase-4: Data Analysis/Design of the Enhanced Concept Based Approach 50
3.2.5 Phase-5: Development / Testing of the 50

Enhanced Concept Based Approach using the BM25 Algorithm
3.2.6 Phase-6: Evaluation/Comparison of the Enhanced Concept Based Approach with the existing Approaches 51
3.3 Overview of the Research Framework 52
3.3.1 End users (Professional and Laymen) 53
3.3.2 User Centered Health Information Retrieval System 54
3.3.3 Search Engines 54
3.3.4 Free Online Medical Datasets (Medline + 7 Others) 54
3.3.5 Medical Terms Control Module 55
3.3.6 Professional End Users (Clinical and Medical Personnel) 55
3.3.7 Unprofessional End Users (Laymen and their Relatives) 55
3.3.8 Vocabulary Controlled Module 56
3.3.9 Readability Module 56
3.4 Experimental Setup 57
3.5 Analysis of Results 57
3.5.1 Performance Analysis Using the Metrics (P@40, NDCG@40, and MAP 58
3.5.2 Statistical Analysis of Performance Evaluation 59
3.6 Chapter Summary 60

CHAPTER 4 THE PROPOSED ENHANCED CONCEPT BASED APPROACH 62

4.1 Introduction 62
4.2 The Model for the Enhanced Concept Based Approach 62
4.3 The Enhanced Concept Based Approach for User Centered Health Information Retrieval 64
4.4 Insightful Comparison of the Existing Concept Based Approach with the Enhanced Concept Based Approach 68
4.5 The Implementation of the Enhanced Concept Based Approach for User Centered Health 69

Information Retrieval
4.5.1 Addressing Readability Issues encountered by laymen patients and their relatives 71
4.5.2 Addressing Vocabulary Mismatched Issues between laymen's queries and expert vocabularies during query expansion 74
4.5.3 Addressing Presentation Issues between Medical Experts and laymen Patients 80
4.6 Chapter Summary 86

CHAPTER 5 SIMULATION RESULTS AND DISCUSSION 87

5.1 Introduction 87
5.2 Simulation Results Based on Accuracy Rate Using Standard Dataset 88
5.2.1 Simulation results for Medline dataset 88
5.2.2 Simulation results for Unified Medical Language System (UMLS) dataset 88
5.2.3 Simulation results for Medical Subject Heading (MeSH) dataset 89
5.2.4 Simulation results for Metamap dataset 90
5.2.5 Simulation results for Metathesaurus dataset 90
5.2.6 Simulation results for Diagnosia 7 dataset 91
5.3.7 Simulation for Khresmoi Project 6 dataset 91
5.3.8 Simulation results for Genetic Home Reference dataset 92
5.3 Validation of System Performance of the Enhanced Concept Based Approach with the Existing Approaches 93
5.4 Evaluation of Simulation Results in Respect to Addressing Readability Issues 95
5.5 Evaluation of Simulation Results in Respect to Addressing Vocabulary Mismatched Issues 98
5.6 Evaluation of Simulation Results in Respect to Addressing Presentation Issues 107
5.7 Statistical Analysis of Performance Evaluation 113
5.7.1 The Analysis of Variance (ANOVA) Test 113
5.8 Chapter Summary 121

CHAPTER 6 CONCLUSION AND FUTURE WORK **123**

6.1 Introduction 123

6.2 Research Summary 122

6.3 Research Contribution 126

6.4 Future Work 127

6.5 Concluding Remarks 128

REFERENCES **129**

VITAE **143**

LIST OF TABLES

2.1	A Summary of reviewed related literature on information retrieval algorithms	34
2.2	A Summary of reviewed related literature on information retrieval features used	36
2.3	A Summary of reviewed related literature on information retrieval models	37
3.1	Characteristics of the datasets	48
4.1	Comparison between Existing Concept Based Approach and Enhanced Concept Based Approach	68
5.1	Simulation Results for Medline Dataset	88
5.2	Simulation Results for UMLS Dataset	89
5.3	Simulation Results for MeSH Dataset	89
5.4	Simulation Results Rate for Metamap Dataset	90
5.5	Simulation Results for Metathesaurus Dataset	91
5.6	Simulation Results for Diagnosia 7 Dataset	91
5.7	Simulation Results for Khresmoi Project 6 Dataset	92
5.8	Simulation Results for Genetic Home Reference Dataset	93
5.9	Accuracy Rate Percentages and Performance Metrics Scores in respect to addressing Readability Issues Using Medline Dataset	96
5.10	Accuracy Rate Percentages and Performance Metrics Scores in respect to addressing Readability Issues Using UMLS Dataset	96
5.11	Accuracy Rate Percentages and Performance Metrics Scores in respect to addressing Readability Issues Using MeSH Dataset	96

5.12	Accuracy Rate Percentages and Performance Metrics Scores in respect to addressing Readability Issues Using Metamap Dataset	96
5.13	Accuracy Rate Percentages and Performance Metrics Scores in respect to addressing Readability Issues Using Metathesaurus Dataset	97
5.14	Accuracy Rate Percentages and Performance Metrics Scores in respect to addressing Readability Issues Using Diagnosia 7 Dataset	97
5.15	Accuracy Rate Percentages and Performance Metrics Scores in respect to addressing Readability Issues Using Khresmoi Project 6 Dataset	97
5.16	Accuracy Rate Percentages and Performance Metrics Scores in respect to addressing Readability Issues Using Genetic Home Reference Dataset	97
5.17	Simulation Scores of ECBA and Existing Approaches CBA, QLM, LSI In Respect to maxSim and idf Values Using MeSH Dataset	101
5.18	Simulation Scores of ECBA and Existing Approaches CBA, QLM, LSI In Respect to maxSim and idf Values Using Khresmoi Project 6 Dataset	101
5.19	Simulation Scores of ECBA and Existing Approaches CBA, QLM, LSI In Respect to maxSim and idf Values Using UMLS Dataset	102
5.20	Simulation Scores of ECBA and Existing Approaches CBA, QLM, LSI In Respect to maxSim and idf Values Using Metamap Dataset	103
5.21	Simulation Scores of ECBA and Existing Approaches CBA, QLM, LSI In Respect to maxSim and idf Values Using Medline Dataset	103
5.22	Simulation Scores of ECBA and Existing Approaches CBA, QLM, LSI In Respect to maxSim and idf Values Using Metathesaurus Dataset	104

5.23	Simulation Scores of ECBA and Existing Approaches CBA, QLM, LSI In Respect to maxSim and idf Values Using Diagnosia 7 Dataset	105
5.24	Simulation Scores of ECBA and Existing Approaches CBA, QLM, LSI In Respect to maxSim and idf Values Using Genetic Home Reference Dataset	105
5.25	Simulation Scores of ECBA and Existing Approaches CBA, QLM, LSI In Respect to maxSim and idf Values Using MeSH Dataset	107
5.26	Simulation Scores of ECBA and Existing Approaches CBA, QLM, LSI In Respect to maxSim and idf Values Using Metamap Dataset	108
5.27	Simulation Scores of ECBA and Existing Approaches CBA, QLM, LSI In Respect to maxSim and idf Values Using Diagnosia 7 Dataset	109
5.28	Simulation Scores of ECBA and Existing Approaches CBA, QLM, LSI In Respect to maxSim and idf Values Using Khresmoi Project 6 Dataset	109
5.29	Simulation Scores of ECBA and Existing Approaches CBA, QLM, LSI In Respect to maxSim and idf Values Using Medline Dataset	110
5.30	Simulation Scores of ECBA and Existing Approaches CBA, QLM, LSI In Respect to maxSim and idf Values Using UMLS Dataset	110
5.31	Simulation Scores of ECBA and Existing Approaches CBA, QLM, LSI In Respect to maxSim and idf Values Using Metathesaurus Dataset	111
5.32	Simulation Scores of ECBA and Existing Approaches CBA, QLM, LSI In Respect to maxSim and idf Values Using Genetic Home Reference Dataset	112
5.33	Mean and Standard Deviation Comparison of ECBA and Existing Approaches CBA, QLM, LSI Using Medline Dataset	114

5.34	A One Way Analysis of Variance For ECBA and Existing Approaches CBA, QLM, LSI Using Medline Dataset	114
5.35	Mean and Standard Deviation Comparison of ECBA and Existing Approaches CBA, QLM, LSI Using UMLS Dataset	115
5.36	A One Way Analysis of Variance For ECBA and Existing Approaches CBA, QLM, LSI Using UMLS Dataset	115
5.37	Mean and Standard Deviation Comparison of ECBA and Existing Approaches CBA, QLM, LSI Using MeSH Dataset	116
5.38	A One Way Analysis of Variance For ECBA and Existing Approaches CBA, QLM, LSI Using MeSH Dataset	116
5.39	Mean and Standard Deviation Comparison of ECBA and Existing Approaches CBA, QLM, LSI Using Metamap Dataset	117
5.40	A One Way Analysis of Variance For ECBA and Existing Approaches CBA, QLM, LSI Using Metamap Dataset	117
5.41	Mean and Standard Deviation Comparison of ECBA and Existing Approaches CBA, QLM, LSI Using Metathesaurus Dataset	117
5.42	A One Way Analysis of Variance For ECBA and Existing Approaches CBA, QLM, LSI Using Meta-thesaurus Dataset	118
5.43	Mean and Standard Deviation Comparison of ECBA and Existing Approaches CBA, QLM, LSI Using Diagnosia Dataset	118
5.44	A One Way Analysis of Variance For ECBA and Existing Approaches CBA, QLM, LSI Using Diagnosia Dataset	118

5.45	Mean and Standard Deviation Comparison of ECBA and Existing Approaches CBA, QLM, LSI Using Khresmoi Project 6 Dataset	119
5.46	A One Way Analysis of Variance For ECBA and Existing Approaches CBA, QLM, LSI Using Khresmoi Project 6 Dataset	119
5.47	Mean and Standard Deviation Comparison of ECBA and Existing Approaches CBA, QLM, LSI Using Genetic Home Reference Dataset	120
5.48	A One Way Analysis of Variance For ECBA and Existing Approaches CBA, QLM, LSI Using Genetic Home Reference Dataset	120

LIST OF FIGURES

2.1	Existing Concept Based Approach	14
3.1	Research Process (Flowchart)	43
3.2	Research Framework	53
4.1	The Model for the Enhanced Concept Based Approach	63
4.2	The Enhanced Concept Based Approach for User Centered Health Information Retrieval	66
4.3	The Web Based Homepage for the Enhanced Concept Based Approach	70
4.4	The Function for Creating Medical Discharge Report in Layman's Form	72
4.5	The Sample Output of Medical Discharge Report in Layman's Form	72
4.6	The Function for Generating Medical Search Query in Layman's Form	73
4.7	The Sample Output of Medical Search Query Result in Layman's Form	74
4.8	The Medical Terms Control Module	75
4.9	The Vocabulary Control Module	76
4.10	The Sample Output generated by the Enhanced Concept Based Approach in respect to addressing Vocabulary Mismatched Issues	77
4.11	The Sample Output generated by the Concept Based Approach in respect to addressing Vocabulary Mismatched Issues	78

4.12	The Sample Output generated by the Enhanced Concept Based Approach in respect to addressing Vocabulary Mismatched Issues	79
4.13	The Sample Output generated by the Concept Based Approach in respect to addressing Vocabulary Mismatched Issues	80
4.14	The Function for Creating Medical Discharge Reports in Expert Form	81
4.15	The Interface View of Medical Discharge Reports in Expert Form	82
4.16	The Interface View of Medical Discharge Reports in Layman's Form	83
4.17	The Function for Creating Medical Search Query in Expert Form	84
4.18	The Interface View of Medical Search Query in Expert Form	84
4.19	The Interface View of Medical Search Query in Layman's Form	85

LIST OF ALGORITHMS

2.1	The Existing Concept Based Approach Algorithm	14
4.2	The Enhanced Concept Based Approach For User Centered Health Information Retrieval Algorithm	66
4.4	The Function for Creating Medical Discharge Document In Layman's Form Algorithm	72
4.6	The Function for Creating Medical Search Query in Layman's Form Algorithm	73
4.8	The Medical Terms Controle Module Algorithm	75
4.9	The Vocabulary Controlled Module Algorithm	76
4.14	The Function for Creating Medical Discharge Document in Expert Form Algorithm	81
4.17	The Function for Creating Medical Search Query in Expert Form Algorithm	84

LIST OF SYMBOLS AND ABBREVIATIONS

η	-	Learning rate
α	-	Momentum coefficient
e	-	Exponent
$f(x)$	-	Function of x
f	-	The squashing or activation function of the Processing unit
$x < 0$	-	x is less than 0
$x > 1$	-	x is greater than 1
$x \geq 0$	-	x is greater than or equal to 0
$-1 \leq x \leq 1$	-	x is greater than or equal to -1 and is less than or
A	-	Relevant Information
B	-	Non Relevant Information
LM	-	Language Modeling Method
VSM	-	Vector space Model
K	-	Denotes the Expansion Terms
CBA	-	Concept Based Approach
MAP	-	Mean Average Precision
NDCG	-	Normalized Discounted Cumulative Gains
P	-	Precision
UMLS	-	Unified Medical Language System
MeSH	-	Medical Subject Heading
N	-	Number of concept terms derived from the original query
SQ	-	Search Query

CCn	-	Represents the nth concept term
#	-	Represent Space Character
SearchQR	-	Search Query
HTML	-	Hyper Text Markup Language
CSS	-	Cascading Style Sheet
LMT	-	Labeled Most Specific Terms
MST	-	Most Specific Terms
ET	-	Denotes the Expansion Terms
ETk	-	Denotes the kth Expansion Term
NA	-	New Algorithm
T	-	Number of Terms
D	-	Denotes Documents
REC-NO	-	Denotes Record Number
DCG	-	Is used to emphasize highly relevant document
Qj	-	Is the number of highly relevant documents
LSI	-	Denotes Latent Semantic Indexing
QLM	-	Denotes Query Likelihood Model
ECBA	-	Denotes Enhanced Concept Based Approach
F1	-	Represents function that create and display medical discharge reports in expert forms.
F2	-	Represents function that create and display medical search queries results in expert forms.
DRI	-	Represents the addition of the functions F1 and F2
F3	-	Represents function that generate and display medical discharge reports in layman's forms.
F4	-	Represents function that generate and display medical search queries results in layman's forms.
DR2	-	Represents the addition of the functions F3 and F4
MTCM	-	Represents the medical terms control module
VCM	-	Represents the vocabulary controlled module
RM	-	Represents the readability module
T_1, T_2….T_n	-	Represents set of terms in a sequence
P(T/D)	-	Represents probability of a term T, given the document D.

T1	-	Represents text contained in the input segment /displayed results in medical expert form
T2	-	Represents text contained in the output segment /displayed results in layman's form
SVD	-	Represents Singular Value Decomposition Function
M	-	Represents complex matrix
U	-	Is an $\boldsymbol{m} \times \boldsymbol{m}$ real or complex unitary matrix
V	-	Is an $\boldsymbol{n} \times \boldsymbol{m}$ real or complex unitary matrix
$\sum$	-	Is an $\boldsymbol{m} \times \boldsymbol{n}$ rectangular diagonal matrix with non-negative real numbers on the diaginal
MM*	-	Represents Left Singular Vectors
M*M	-	Represents Right Singular Vectors
ANOVA	-	Analysis of Variance
HSD	-	Highly Significant Difference

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A	Table A.1: Gantt Chart of Research Activities	142
B	My Vitae	144

LIST OF PUBLICATIONS

Journals:

(i) **Kontagora I.U. & Hamid I.R.A,** "An Enhanced Algorithm for User Centered Health Information Retrieval to Address Readability Issues" Being a Publication Indexed by ISI, SCOPUS, SEMANTIC SCHOLAR, INSPEC DIRECT, WORLD CAT, ULRICHS WEB, EBSCO HOST and SERIALS SOLUTIONS & Published in International Journal of Advanced Computer Science and Applications (IJACSA), 10(3), pp 439 – 446, 2019. http://dx.doi.org/10.14569/IJACSA.2019.0100356.

(ii) **Kontagora I.U., Hamid I.R.A, Omar N.A,** "An Enhanced Concept Based Approach for User Centered Health Information Retrieval to Address Presentation Issues" Being a Publication Indexed by ISI, SCOPUS, SEMANTIC SCHOLAR, INSPEC DIRECT, WORLD CAT, ULRICHS WEB, EBSCO HOST and SERIALS SOLUTIONS & Published in International Journal of Advanced Computer Science and Applications (IJACSA), 10(1), pp 232 – 242, 2019. http://dx.doi.org/10.14569/IJACSA.2019.0100131

(iii) **Kontagora I.U., Hamid I.R.A, Omar N.A,** "An Enhanced Concept Based Approach for User Centered Health Information Retrieval to Address Vocabulary Mismatched Issues" Being a Publication **(Q4: IF 0.18)** Indexed by SCOPUS, DOAJ, GOGGLE SCHOLAR, DRJI, getCITED (USA), ProQuest (USA),OPEN J-Gate, DBLB and Science Central & Accepted and In Press International Journal of Engineering and Technology(UAE). 2019 Edition.

Proceedings:

(i) **Kontagora I.U. & Hamid I.R.A,** "Comparative Studies of Information Retrieval Approaches in User Centered Health Information System" Being a Publication Indexed by ISI, Scopus, Goggle Scholar, SpringerLink, Ulrich's, DBLP, Ei-Compendex, Zentralblatt and Metapress & Published in Ghazali R., Deris M., Nawi N., Abawajy J. (eds), Recent Advances on Soft Computing and Data Mining. SCDM 2018. Advances in Intelligent Systems and Computing, vol 700 Pp 171 – 180. Springer, Cham". 2018. https://doi.org/10.1007/978-3-319-72550-5_17

REFERENCES

Appri A., Carian E. K., Johnson M. E., and Anthony D. L., "Medication administration quality and health information technology," a national study of US hospitals, J. Am. Med. Inform. Assoc. 19 (3) (2012) 360–367, 2012.

Aceto G., Persico V. and Pescapé A., "The role of Information and Communication Technologies in healthcare: taxonomies, perspectives, and challenges," In Journal of Network and Computer Applications, 2018.

Alyami M. A, "Toward patient-centered personal health records systems to promote evidence-based decision-making and information sharing, 2018.

Almasri M., Berrut C and Chevallet J. P, "A comparison of deep learning based query expansion with pseudo-relevance feedback and mutual information," In European conference on information retrieval (pp. 709-715). Springer, Cham, 2016.

Ayatollahi H., Bath P. A and Goodacre S., "Factors influencing the use of IT in the emergency department: a qualitative study," In Health Inf. J. 16 (3) (2010) 189–200, 2010.

Aravazhi R and Chidambaram M., "A Secure Electronic Medical Information Retrieval System in EMIRS," In International Journal of Applied Engineering Research, 12(12), 3500-3506, 2017.

Anagnostou V., Smith K. N, Forde P. M, Niknafs N., Bhattacharya R., White J and Hruban C., "Evolution of neoantigen landscape during immune checkpoint blockade in non–small cell lung cancer," In *Cancer discovery*, *7*(3), 264-276, 2017.

Brahmandam S., Holland W. C, Mangipudi S. A, Braz V. A., Medlin R. P and Hunold K. M "Willingness and ability of older adults in the emergency department to provide clinical information using a tablet computer," In J. Am. Geriatr. Soc. 64 (11) (2016) 2362–2367, 2016.

Beuscart-Zéphir M. C, Elkin P., Pelayo S and Beuscart R., "The human factors engineering approach to biomedical informatics projects: state of the art, results, benefits and challenges," In Yearb. Med. Inf. 10 (2007) 9–27, 2007.

Bell D., Gachuhi N and Assefi N., "Dynamic Clinical Algorithms: Digital Technology Can Transform Health Care Decision-Making," In the American journal of tropical medicine and hygiene. 98(1): pp 9-14. 2018.

Burstein F., Fisher J., McKemmish S., Manaszewicz R. and Malhotra P, "User Centred Quality Health Information Provision: Benefits and Challenges," In System Sciences, 2012. In Proceedings of the 38th Annual Hawaii International Conference, pp.138c, 138, 2012.

Burden Scales for Cancer Clinical Trial Participation," In Journal of Empirical Research on Human Research Ethics, 1556264618764730, (2018).

Bardhan I. R. and Thouin M. F., "Health information technology and its impact on the quality and cost of healthcare delivery, Decis," In Support Syst. 55 (2) (2013) 438– 449, 2013.

Banawan K. and Ulukus S., "The capacity of private information retrieval from coded databases," In IEEE Transactions on Information Theory, 64(3), pp.1945-1956. 2018.

Chiuchisan I. U. L. I. A. N. A. and Geman O. A. N. A., "An approach of a decision support and home monitoring system for patients with neurological disorders using internet of things concepts," In WSEAS Transations on Systems, 13, 460- 469, 2014.

Choi S and Choi J., "Exploring effective information retrieval technique for the medical web documents: SNUMedinfo," In CLEFeHealth2014 Task 3. 2013.

Chen B., Tsutsui S., Ding Y. and Ma F., "Understanding the topic evolution in a scientific domain, "An exploratory study for the field of information retrieval. Journal of Informetrics, 11(4), 1175-1189, 2017.

Cresswell K., Morrison Z., Crowe S., Robertson A. A and Sheikh A., "Anything but engaged: user involvement in the context of a national electronic health record implementation," In Inform. Prim. Care 19 (4) (2011) 191–206, 2011.

Clarke M. A, Moore J. L, Steege L. M, Koopman R. J, Belden J. L, Canfield S. M, and Kim M. S, " Health information needs, sources, and barriers of primary care patients to achieve patient-centered care: A literature review," In Health informatics journal, *22*(4), 992-1016, 2016.

Chandraratne N. K, Pathirathna K. G, Harrison C and Siriwardena A. N, "A comparison of policies and guidelines related to multimorbidity in the UK, Australia and Sri Lanka," In Australian journal of general practice, *47*(1-2), 15-19, 2018.

Claveau V., Hamon T., Grabar N and Maguer S. L, "RePaLi participation to CLEF eHealth IR challenge 2014: leveraging term variation," In Proceedings of the ShARe/CLEF eHealth Evaluation Lab, 2014.

Choi S and Choi J., "Exploring effective information retrieval technique for the medical web documents: SNUMedinfo at CLEF eHealth 2014 Task 3," In Proceedings of the ShARe/CLEF eHealth Evaluation Lab, 2014.

Cismondi F., Fialho A. S., Vieira S. M., Reti S. R., Sousa J. M., and Finkelstein S. N., Missing data in medical databases: Impute, delete or classify?" Artificial intelligence in medicine, vol. 58, no. 1, pp. 63-72, 2013.

Chengxiang Z and John L., "Model-based feedback in the language modeling approach to information retrieval," In Proceedings of the Tenth ACM International Conference on Information and Knowledge Management (CIKM'01), pages 403–410, 2001.

Djenouri Y., Belhadi A and Belkebir R., "Bees swarm optimization guided by data mining techniques for document information retrieval. In Expert Systems with Applications. 94: pp 126-136. 2018.

Deo A., Gangrade J. and Gangrade S., "A SURVEY PAPER ON INFORMATION RETRIEVAL SYSTEM," In International Journal of Advanced Research in Computer Science. 9(1), 2018.

Donald H. K., Bruce Croft W., David J. H. and Justin Z. (eds.), "Proceedings of the Association for Computing Machinery," 2001.

Djoerd H. and Franciska De J., "Disambiguation strategies for cross-language information retrieval," In Lecture Notes in Computer Science. Springer-Verlag, pages 274– 293, 2015.

Drame K., Mougin F and Diallo G., "Query expansion using external resources for improving information retrieval in the biomedical domain," In Proceedings of the ShARe/CLEF eHealth Evaluation Lab, 2014.

Dybkjaer L., Hemsen H and Minker W. (Eds.), "An Overview of evaluation methods. Evaluation of Text and Speech Systems, "In TREC Ad-hoc Information

Retrieval and TREC Question Answering. Springer, Dordrecht, the Nethe, 2007.

David M. B, Andrew Y. N and Michael I. J, "Latent Dirichlet allocation," In Journal of Machine Learning Research, 2003.

Deshpande P., Rasin A., Brown E., Furst J., Raicu D., Montner S and Armato-III S., "An Integrated Database and Smart Search Tool for Medical Knowledge Extraction from Radiology Teaching Files," In Medical Informatics and Healthcare, pp. 10-18, 2017.

Egghe L., "Power laws in the information production process," In Lotkaian Informetrics, 2005.

Fuad A. and Hsu C. Y., "High rate EHR adoption in Korea and health IT rise in Asia," In *Journal Med. Inf*, 81(9), 649–650, 2012.

Fuchs C and Groh G., "Data sharing sensitivity, relevance, and social traits in social information retrieval," In International Journal of Social Computing and Cyber- Physical Systems, 2(1), 1-34, 2017.

Farhangfar A., Kugan L., and Dy J., "Impact of imputation of missing values on classification error for discrete data," Pattern Recognition, vol. 41, no. 12, pp. 3692-3705, 2008.

Fishman J. M., Greenberg P., Bagga M. B., Casarett D and Propert K., "Increasing Information Dissemination in Cancer Communication: Effects of Using "Palliative", "Supportive," or "Hospice" Care Terminology," In Journal *of palliative medicine*, 2018.

Fox S., "Health topics: 80% of internet users look for health information online," In Technical report, Pew Research Center, (February 2011)

Ford D. T., "Are electronic health records the future of dental practice?," In J. Calif. Dent. Assoc. 43 (5) (2015) 238. 2015.

Ford E. W., Menachemi N., Peterson L. T and Huerta T. R., "Resistance is futile: but it is slowing the pace of EHR adoption nonetheless," In J. Am. Med. Inform. Assoc. 16 (3) (2009) 274–281, 2009.

Ghoulam A., Barigou F., Belalem G. and Meziane F., "Query Expansion Using Medical Information Extraction for Improving Information Retrieval in French Medical Domain," In International Journal of Intelligent Information Technologies (IJIIT), 14(3), 1-17, 2018.

Giunti G., Giunta D. H, Guisado-Fernandez E., Bender J. L and Fernandez-Luque L., "A biopsy of Breast Cancer mobile applications: state of the practice review," In International journal of medical informatics, 110, 1-9, 2018.

Griffith M., Spies N. C., Krysiak K., McMichael J. F., Coffman A. C., Danos A. M., and Barnell E. K., "CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer," In Nature genetics, 49(2), 170, (2017).

Goeuriot L., Kelly L., Li W., Palotti J., Pecina P., Zuccon G., Hanbury A., Jones G., Mueller H., "ShARe/CLEF eHealth Evaluation Lab 2013, Task 2: User centered health information retrieval," In Proceedings of CLEF 2013 (2013).

Garcia S., ,Luengo J. and Herrera F., "Data preprocessing in data mining. Springer, 2015

Goeuriot L., Kelly L., Li W., Palotti J., Pecina P., Zuccon G., A. Hanbury, G. Jones, H. Mueller, "ShARe/CLEF eHealth Evaluation Lab 2014, Task 3: User centered health information retrieval," In Proceedings of CLEF 2014 (2014).

Gray K and Gilbert C., "Digital Health Research Methods and Tools: Suggestions and Selected Resources for Researchers," In *Advances in Biomedical Informatics* (pp. 5-34). Springer, Cham, 2018.

Gely P. B., Amy N. L and Valeriy A. N., "The life and work of A.A. Markov," In Linear Algebra and its Applications, 2004.

Garcelon N., Neuraz A., Salomon R., Faour H., Benoit V., Delapalme A and Rance B.,"A clinician friendly data warehouse oriented toward narrative reports: Dr. Warehouse," In Journal of biomedical informatics, 80, 52-63, (2018).

Greysen S. R., Khanna R. R., Jacolbia R., Lee H. M., Auerbach A. D., "Tablet computers for hospitalized patients: a pilot study to improve inpatient engagement," In J. Hosp. Med. 9 (6) (2014) 396–399, 2014.

Guido Z., Joao P., Lorraine G., Liadh K., Mihai L., Pavel P., Henning M., Julie B and Anthony D., "The IR Task at the CLEF eHealth Evaluation Lab 2016," User-Centered Health Information Retrieval, 2016.

Hiemstra D., "Using language models for information retrieval," Taaluitgeverij Neslia Paniculata, 2011.

Hou H., Kretschmer H. and Liu Z., "The structure of Collaboration networks in Scientometrics," Scientometrics, 75(2), 189–202. Jackendoff, R. S. (1992). Languages of the mind. Cambridge, MA: MIT Press, 2008.

Halbert C. H and Harrison B. W., "Genetic counseling among minority populations in the era of precision medicine," In American Journal of Medical Genetics Part C: Seminars in Medical Genetics, 2018.

Huang, Norden Eh., "Hilbert-Huang transform and its applications," Vol. 16. World Scientific, 2014.

Hanauer D. A, Wu D. T., Yang T. L, Mei Q., Murkowski-Steffy K. B., Vydiswaran V. V and Zheng K., "Development and empirical user-centered evaluation of semantically-based query recommendation for an electronic health record search engine," In Journal of biomedical informatics, 67, 1-10, 2017.

Henry E., Silva A., Tarlov E., Czerlanis C., Bernard M., Chauhan C and Stewart G., "Delivering coordinated cancer care by building transactive memory in a team of teams," In Journal of oncology practice, 12(11), 992-999, (2016).

Irizarry T., Vito A. D and Dabbs C. R., "Curran, Patient portals and patient engagement: a state of the science review," In Journal of Med. Internet Res. 17 (6) e148, 2015

James A., Jay A., Nicholas B. and Chris B., "Challenges in Information Retrieval and Language Modeling," In SIGIR Forum 37(1), 2003.

Joho H., Cavedon L., Arguello J., Shokouhi M. and Radlinski F., "2018, February. CAIR'17: First International Workshop on Conversational Approaches to Information Retrieval at SIGIR 2017," In *ACM SIGIR Forum*, Vol. 51, No. 3, pp. 114-12). ACM. 2018.

Jaruskulchai C., and Thesprasith O., "Csku gprf-qe for medical topic web retrieval," In proceedings of the ShaRe/CLEF eHealth Evaluation Lab. 2015.

Jinxi X and Ralph W., "A probabilistic approach to term translation for cross-lingual retrieval," In W. Bruce Croft and John Lafferty (eds.) Language Modeling for Information Retrieval, Kluwer Academic Publishers, pages 125– 140, 2003.

Joao P., Lorraine G., Liadh K., Wei L., Pavel P. G., "ShARe/CLEF eHealth Evaluation Lab 2015, Task 3," User-centred health information retrieval, 2015.

Jiménez-Zafra S. M., Martín-Valdivia M. T., Molina-González M. D and Ureña-López L. A., "How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for medical domain," In Artificial Intelligence in Medicine, 2018.

Kellermann A. L. and Jones S. S., "What it will take to achieve the as-yet-unfulfilled promises of health information technology, Health Aff," In (Millwood) 32 (1) (2013) 63–68, 2013.

Koopman B., Zuccon G., Bruza P., Sitbon L. and Lawley M., "An evaluation of corpus-driven measures of medical concept similarity for information retrieval," In Proceedings of CIKM, 2012.

Kushniruk A. and Nøhr C., "Participatory design, user involvement and health IT evaluation," In Stud. Health Technol. Inform. 222 (2016) 139–151, 2016.

Kruse C. S., Argueta D. A., Lopez L. and Nair A., "Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review," In J. Med. Internet Res. 17 (2) e40 (2015).

Kruse C. S., Bolton K. and Freriks G., "The effect of patient portals on quality outcomes and its implications to meaningful use: asystematic review," In Journal of Med. Internet Res. 17 (2) (2015) e44, 2015.

Kalpathy G. J., M• uller H., Bedrick S., Eggel I., Herrera de A. G. S., Tsikrika T., "The CLEF 2013 medical image retrieval and classification tasks," In: Working Notes of CLEF 2013, (Cross Language Evaluation Forum). (2013)

Kwon H., Kim J and Park Y., "Applying LSA text mining technique in envisioning social impacts of emerging technologies: The case of drone technology," In Technovation, *60*, 15- 28, 2017.

Kitajima M., Kariya N., Takagi H and Zhang Y., "Evaluation of website usability using Markov chains and latent semantic analysis," In IEICE Transactions on Communications, E88B(4), 1467–1475, 2005.

King J., Patel V., Jamoom E. W., Furukawa M. F., "Clinical benefits of electronic health record use: national findings," In HealthServ.Res.49 (1pt.2) (2014)392–404, 2014.

Kang H., "The prevention and handling of the missing data," Korean journal of anesthesiology, vol. 64, no. 5, pp. 402-406, 2013

Kelly L., Goeuriot L., Suominen H., Schrek T., Leroy G., Mowery D. L., Velupillai S., Chapman W. W., Martinez D., Zuccon G and Palotti J., "Overview of the ShARe/CLEF eHealth Evaluation Lab 2014," In Proceedings of CLEF 2014. Springer (2014).

Kelly L., Goeuriot L., Suominen H., Névéol A., Palotti J and Zuccon G. , "Overview of the CLEF eHealth evaluation lab 2016," In International Conference of the

Cross-Language Evaluation Forum for European Languages (pp.255-266). Springer, Cham, 2016.

Krishnamurthy S and Akila V., "Information Retrieval Models: Trends and Techniques," In *Web Semantics for Textual and Visual Information Retrieval* (pp. 17-42). IGI Global, 2006.

Ksentini N., Tmar M and Gargouri F., "Miracle at CLEF 2014: eHealth information retrieval task," In: Proceedings of the ShARe/CLEF eHealth Evaluation Lab, 2014.

Koswatte S., McDougall K and Liu X., "Relevance assessment of crowdsourced data (CSD) using semantics and geographic information retrieval (GIR) techniques," In ISPRS International Journal of Geo-Information, *7*(7), p.256. 2018.

Lablans M., Schmidt E. E and Ückert F., "An Architecture for Translational Cancer Research As Exemplified by the German Cancer Consortium," In *JCO Clinical Cancer Informatics, 1*, 1-8, 2018.

Liui X., Shen W., Nie J.Y and Liu X. , "An investigation of the effectiveness of concept based approach in medical information retrieval GRIUM @ CLEF2015eHealthTask 3," In Proceedings of the ShARe/CLEF eHealth Evaluation Lab, 2015.

Liu Y., Shi J and Chen Y., "Patient-centered and experience-aware mining for effective adverse drug reaction discovery in online health forums," In Journal of the Association for Information Science and Technology, 69(2), 215-228, (2018).

Larose D. T., "Discovering knowledge in data mining: an introduction to data. John Wiley Sons, 2014.

Leech, Nancy L., Karen C. Barrett, and George A. Morgan. *IBM SPSS* for intermediate statistics: Use and interpretation. Routledge, 2014.

Leydesdorff L., '"Betweenness centrality" as an indicator of the "interdisciplinary" of scientific journals,". In Journal of the American Society for Information Science and Technology, 58, 1303–1319, 2007.

Leveling J., Goeuriot L., Kelly L., Jones G.J.F, "DCU@TRECMed 2014: Using ad-hoc baselines for domain-specific retrieval," In: Proceedings of TREC 2014, NIST, (2014)

Manogaran G. and Lopez D., "A survey of big data architectures and machine learning algorithms in healthcare," In International Journal of Biomedical Engineering and Technology. 25(2-4): pp 182-211. 2017.

Menad H. and Amine A., "Bio-Inspired Algorithms for Medical Data Analysis," In Handbook of Research on Biomimicry in Information Retrieval and Knowledge Management 2018. pp. 251-275. IGI Global. 2018.

Martijn S and Wessel K., "Language models for topic tracking," In W. Bruce Croft and John Lafferty (eds.) Language Modeling for Information Retrieval, Kluwer Academic Publishers, pages 95–124, 2003.

McRoy S., Mojarad M. R., Wang Y., Ruddy K. J, Haddad T. C and Liu H., "Assessing Unmet Information Needs of Breast Cancer Survivors: Exploratory Study of Online Health Forums Using Text Classification and Retrieval," In JMIR cancer. 2018. (1). 2018.

Maiella S., Olry A., Hanauer M., Lanneau V., Lourghi H., Donadille B and Parkinson H., "Harmonising phenomics information for a better interoperability in the rare disease field," In European journal of medical genetics, 2018.

Majumder P., Iyer G., Shah K and Thakkar H., "Team IRLabDAIICT at ShARe/CLEF eHealth 2015 Task 3: User-centered Information Retrieval system for Clinical Documents," In: Proceedings of the ShARe/CLEF eHealth Evaluation Lab, 2015.

Newman M. E. J., "Who is the best connected scientist? A study of scientific coauthorship networks," In E. Ben-Naim, H. Frauenfelder, & Z. Toroczkai (Eds.), Complex networks (pp. 337–370). Berlin: Springer. From http://www-personal.umich.edu /~mejn/ papers/ cnlspre pdf Retrieved 11.08.09, 2004.

Nguyen G. H, Tamine L., Soulier L. and Souf N., "Learning Concept-Driven Document Embeddings for Medical Information Search," In Conference on Artificial Intelligence in Medicine in Europe (pp. 160-170). Springer, Cham, 2017.

Nguyen L., Bellucci E and Nguyen L. T., "Electronic health records implementation: an evaluation of information system impact and contingency factors," In International Journal of Med. Inf. 83 (11) (2014) 779–796, 2014.

Oh H. S. and Jung Y., "A multiple-stage approach to re-ranking clinical documents," In Proceedings of the ShARe/CLEF eHealth Evaluation Lab, 2014.

Ornuma T Chuleerat and J., "CSKU GPRF-QE for Medical Topic Web Retrieval," Department of Computer Science, Faculty of Science, Kasetsart University, Bangkok, Thailand ornuma.thesprasith@gmail.com and fscichj@ku.ac.th. (2015).

Ozturkmenoglu O., Alpkocak A and Kilinc D., "Demir at CLEF eHealth: The effects of selective query expansion to information retrieval," In proceedings of the ShARe/CLEF eHealth Evaluation Lab, 2014.

Paula A. A. D, Pires D. F., Filho P. A., de Lemos K. R. V., Barçante E. and Pacheco A. G., "A comparison of accuracy and computational feasibility of two record linkage algorithms in retrieving vital status information from HIV/AIDS patients registered in Brazilian public databases ," In International journal of medical informatics, 114, pp.45-51. 2018

Pallant J.**, "**SPSS survival manual: A step by step guide to data analysis using SPSS", 2011

Patterson A. L., inventor L.L.C, Google and assignee, "Multiple index based information retrieval system," In United States patent application US 15/809,356. 2018.

Pollack A. H., Miller A., Mishra S. R, Pratt W. and Atricians P. D., "leveraging physicians and participatory design to develop novel clinical information tools," In AMIA Annual Symposium Proceedings/AMIA Symposium AMIA Symposium. 2016 (2016) 1030–1039.

Papoutsi E., Reed J. E., Marston C., Lewis R., Majeed A. and Bell D., "Patient and public views about the security and privacy of electronic health records (EHRs) in the UK: results from a mixed methods study," In BMC Med. Inform. Decis. Mak. 15 (1) (2015) 1. 2015

Pinciroli F and Pagliari C., "Understanding the evolving role of the Personal Health Record, Compute," In Biol. Med. 59 (2015) 160–163, (2015).

Pestian J. P, Matykiewicz P., Duch W., Glauser T. A., Kowatch R. A., Grupp-Phelan J. M. and Sorter M., "Cincinnati Children s Hospital Medical Center, 2016," In *Processing text with domain-specific spreading activation methods*. U.S. Patent 9,477,655, (2016).

Richard M. S., Sreenivasa S and Timothy L., "Unsupervised Topic Discovery," In Proceedings of the Language Models for Information Retrieval workshop (LMIR), 2001.

Suominen H., Kelly L., Goeuriot L., Hanlen L., Névéol A., Grouin C. and Zuccon G., "Overview of the CLEF eHealth evaluation lab 2015," In In International Conference of the Cross-Language Evaluation Forum for Languages (pp. 429-443). Springer, Cham. (2015)

Shen W., Nie J. Y., Liu X and Liui X., "An investigation of the effectiveness of concept-based approach in medical information retrieval GRIUM @ CLEF2014eHealthTask 3," In: Proceedings of the ShARe/CLEF eHealth Evaluation Lab, 2014.

Shen W., Nie J. Y., Liu X., Liui X., "An investigation of the effectiveness of concept- based approach in medical information retrieval GRIUM @ CLEF2015 eHealth Task 4," In: Proceedings of the ShARe/CLEF eHealth, 2015.

Thesprasith O and Jaruskulchai C., "Csku gprf-qe for medical topic web retrieval," In Proceedings of the ShARe/CLEF eHealth Evaluation Lab, 2014.

Thakkar H., Iyer G., Shah K. and Majumder P., "Team IRLabDAIICT at ShARe/CLEF eHealth 2014 Task 3: User-centered Information Retrieval system for Clinical Documents," In: Proceedings of the ShARe/CLEF eHealth Evaluation Lab, 2014.

Thakkar H., Iyer G. and Majumder P., "A comparative study of approaches in user-centered health information retrieval," In arXiv preprint arXiv: 1505.01606, (2015).

Voorhees E. M and Tong R. M., "Overview of the TREC 2011 medical records track," In: Proceedings of TREC, NIST, (2011).

Victor L and Bruce-Croft W., "Relevance models in information retrieval," In W. Bruce Croft and John Lafferty (eds.) Language Modeling for Information Retrieval, Kluwer Academic Publishers, pages 11–56, (2003).

Wiley J. A., Rittenhouse D. R., Shortell S. M., Casalino L. P., Ramsay P. P., Bibi S and Alexander J. A., "Managing chronic illness: physician practices increased the use of care management and medical home processes," In Health Affairs, *34*(1), 78-86, (2015).

Wessel K., Thijs W and Djoerd H., "The importance of prior probabilities for entry page search," In Proceedings of the ACM Conference on Research and Development in Information Retrieval (SIGIR'02), 2002.

Winer R. L, Tiro J. A., Miglioretti D. L., Thayer C., Beatty T., Lin J. and Buist D. S., "Rationale and design of the HOME trial: A pragmatic randomized controlled trial of home-based human papillomavirus (HPV) self-sampling for increasing cervical cancer screening uptake and effectiveness in a US healthcare system," In Contemporary clinical trials, 64, pp. 77-87, 2018.

White R and Horvitz E., "Cyberchondria: Studies of the escalation of medical concerns in web search," In Technical report, Microsoft Research, 2009.

Welbie M., Wittink H., Westerman M. J, Topper I., Snoei J and Devillé W. L., "Using Plain Language and Adding Communication Technology to an Existing Health-Related Questionnaire to Help Generate Accurate Information: Qualitative Study," In Journal of medical Internet research, 20(4), (2018).

Xiaoshi Z., Yunqing X., Zhongda X., Sen N., Qinan H and Yaohai H., "Concept-based medical document retrieval: THCIB at CLEF eHealth lab 2013 task 4." In Proceedings of the ShARe/CLEF eHealth Evaluation Lab, 2013.

Yang C., Bhattacharya S. and Srinivasan P., "The University of Iowa at CLEF 2014: eHealth Task 3," In Proceedings of the ShARe/CLEF eHealth Evaluation Lab, 2014.

Zayani S., Ksentini N., Tmar M and Gargouri F., "Miracl at Clef 2018: Consumer Health Search Task," In CEUR Workshop Proceedings: Working Notes of CLEF 2018.

Zaki M. J and Meira jr. W., Data mining and analysis: fundamental concepts and algorithms. Cambridge University Press, NY, USA, 2014.

Zhong X., Xia Y., Xie S., Na Z., Hu Q and Huang Y., "Concept-based Medical Document Retrieval: THCIB at CLEF eHealth Lab 2013 Task 3," In *CLEF* (Working Notes), 2013.